

A multimodal generative AI copilot for human pathology

Year: 2024 | Citations: 292 | Authors: Ming Y. Lu, Bowen Chen, Drew F. K. Williamson, Richard J. Chen, Melissa Zhao

Abstract

Computational pathology^{1,2} has witnessed considerable progress in the development of both task-specific predictive models and task-agnostic self-supervised vision encoders^{3,4}. However, despite the explosive growth of generative artificial intelligence (AI), there have been few studies on building general-purpose multimodal AI assistants and copilots⁵ tailored to pathology. Here we present PathChat, a vision-language generalist AI assistant for human pathology. We built PathChat by adapting a foundational vision encoder for pathology, combining it with a pretrained large language model and fine-tuning the whole system on over 456,000 diverse visual-language instructions consisting of 999,202 question and answer turns. We compare PathChat with several multimodal vision-language AI assistants and GPT-4V, which powers the commercially available multimodal general-purpose AI assistant ChatGPT-4 (ref. 6). PathChat achieved state-of-the-art performance on multiple-choice diagnostic questions from cases with diverse tissue origins and disease models. Furthermore, using open-ended questions and human expert evaluation, we found that overall PathChat produced more accurate and pathologist-preferable responses to diverse queries related to pathology. As an interactive vision-language AI copilot that can flexibly handle both visual and natural language inputs, PathChat may potentially find impactful applications in pathology education, research and human-in-the-loop clinical decision-making. PathChat, a multimodal generative AI copilot for human pathology, has been trained on a large dataset of visual-language instructions to interactively assist users with diverse pathology tasks.